Changing the Narrative Perspective:
From Deictic to Anaphoric Point of View

Mike Chen[a], Razvan Bunescu[b,*]

[a]*School of Electrical Engineering and Computer Science, Ohio University*
[b]*Department of Computer Science, University of North Carolina at Charlotte*

Abstract

We introduce the task of changing the narrative point of view, where characters are assigned a narrative perspective that is different from the one originally used by the writer. The resulting shift in the narrative point of view alters the reading experience and can be used as a tool in fiction writing or to generate types of text ranging from educational to self-help and self-diagnosis. We introduce a benchmark dataset containing a wide range of types of narratives annotated with changes in point of view from deictic (first or second person) to anaphoric (third person) and describe a pipeline for processing raw text that relies on a neural architecture for mention selection. Evaluations on the new benchmark dataset show that the proposed architecture substantially outperforms the baselines by generating mentions that are less ambiguous and more natural.

Keywords: narrative perspective, narrative point of view

1. Introduction and Motivation

The narrative point of view (PoV), also called the narrative perspective, is the position from which the events in a story are observed and communicated (van Peer and Chatman, 2001). There are three types of narrative perspective, mirroring the use of personal pronouns: *first*, *second*, and *third* person. The narrative point of view influences the readers' perspective-taking and their involvement in the characters' thoughts and feelings. In fiction, the most common narrative perspectives are the first and third

---

*Corresponding author.
Email addresses:* `mc277509@ohio.edu` (Mike Chen), `rbunescu@uncc.edu` (Razvan Bunescu)

person. The following sentence, taken from (Herr, 1977), shows a change of the PoV of the main character, a war correspondent in Vietnam, from the original 1st person to 3rd person:

> *That night, probably sleeping, {I} → {Michael} heard the sound of automatic-weapons fire outside. {I} → {He} had no real sense of waking, only of suddenly seeing three cigarettes glowing in the dark.*

The 1st person narration makes the reader feel like they are watching the story unfold through the eye of the journalist, who is the main protagonist. While the writer expresses his feelings about the events he goes through, the focus is on what he witnesses. Changing the PoV to 3rd person, and referring to him as *Michael*, takes some of that focus away and puts it on the journalist.

The second person point of view is more common in poetry, how-to guides, technical writing, and self-help texts, whereas in fiction the bulk of second person literature dates from only the past sixty years (Fludernik, 1993). In one common instantiation of this type of narrative perspective, the reader is incorporated as a character in the story and is addressed by the writer using second person pronouns. The following sentences, taken from (Evans, 2015), use the 2nd person PoV to address the reader as a potential victim of narcissism:

> *It doesn't matter how hard {you} try to do the right thing, the narcissist still finds fault with {you}. In stark contrast, {you} thought {you} could do no wrong with this person.*

While in general the 1st person perspective leads to a more immersive reading experience than a 3rd person perspective, the 2nd person point of view used in the text above is even stronger in that it facilitates an identification of the reader with the protagonist. Fludernik (1993) distinguishes between two major uses of the second person pronoun in narratives: the addressee *you* and the protagonist *you*, where the addressee can be further refined to the self-addressing *you* in interior monologue fiction. She argues that the initial ambiguity between the interpretation of *you* as the reader vs. a protagonist can develop into an increased empathy effect, achieving maximum identification in present tense texts. This can be used, for example, to manipulate readers into an empathetic identification with the protagonist from which it then proves difficult to withdraw

when an important aspect of that character is finally revealed, a technique that forces the reader to become aware of their own prejudice and then proceed to either change (and thus stay consistent with the initial empathetic reading) or withdraw the empathy altogether.

The point of view is important in other modalities as well. In video games, the choice between the first vs. third person visual point of view was shown to moderate psychophysiological responses during play, with avatar choice producing a more pronounced effect in the third-person PoV (where the "camera" was located behind the avatar) (Lim and Reeves, 2009), while the first-person PoV was observed to heighten realism by providing a sense of inhabiting the representational space. The viewpoint use is much more dynamic in film, where it is very common for it to shift from first to third person, as for example in a shot–reverse shot sequence, where we are initially placed to observe events from the position of the film's protagonist, only to be looking back at the protagonist from the viewpoint of another character moments later (Black, 2017).

In the philosophy of mind, the three perspectives are considered to reflect different modes of epistemic access: the first person perspective is *subjective* because it is directed at the epistemic subject's experiences; the third person perspective is *objective* because it provides access to objective knowledge of other entities; the second person perspective is *intersubjective* because "it is an epistemic relation between an epistemic subject and another sentient being's mental states" (Pauen, 2012). Whereas the second person perspective has been relatively less studied (Bohman, 2000), the epistemic differences between the first and third person perspective are well established (knowing ourselves vs. knowing the world around us), with Choifer (2018) proposing that both belong to a *reflective* mode of experience, as opposed to the 'what-it-is-like' *non-reflective*, non-epistemic access to one's conscious experience.

In narratives, the choice of PoV has been shown in multiple empirical studies to have a substantial effect on modulating the reader's engagement, perspective-taking, and overall their immersion in the events and the mental states described in a story. van Lissa et al. (2016) studied how narrative perspective affects readers' responses to a complex, and potentially unreliable character. Their experimental results showed that the

Autobiography ↔ Biography	
1st	◆ When the youngest of *my* sisters saw him in his bloodied condition, *I* clearly recall her bursting out "Let me die in his place!". (Kurosawa, 1983)
3rd	◆ When the youngest of *Akira's* sisters saw her brother in his bloodied condition, *Akira* clearly recalls her bursting out "Let me die in his place!".

Educational ↔ Self-diagnosis	
3rd	◆ *Persons afflicted with the classic form of the illness* have periods of severe depression as well as periods of mania. (Mondimore, 2006)
2nd	◆ *You* have periods of severe depression as well as periods of mania.

Abstract ↔ Related work	
1st	◆ *We* show that the bound is tight, and *we* develop an algorithm that computes the optimal schedule. (Gast et al., 2013)
3rd	◆ *The authors* show that the bound is tight, and *they* develop an algorithm that computes the optimal schedule.

Table 1: Pairs of sentences written from different narrative perspectives. Verbs whose conjugation is changed due to number and person agreement are shown with dotted underlines.

manipulation of narrative perspective influences readers' trust for the character. Salem et al. (2017) ran questionnaire studies to investigate the effect of different textual modes of narration on reading experience. Their work showed that first person narration and psycho-narration (a third person perspective of a character's consciousness) increased the tendency to take over the perspective of the protagonist. Hartung et al. (2016) investigated the effect of personal pronouns on discourse comprehension, with results showing that first person pronouns lead to higher immersion. Using simple event sentences, Brunyé et al. (2009) found that third person pronouns facilitate the construction of a mental model from an observer's perspective, whereas second person pronouns have the strongest effect in making the readers feel they embody an actor's perspective.

1.1. On the utility of changing the narrative perspective

Given the substantial effect that PoV has on the readers' perspective-taking, the manipulation of narrative perspectives can be a valuable tool in fiction writing. Fur-

thermore, in non-fiction, articles written from different perspectives can serve different purposes, some of which are illustrated in Tables 1 and 2. The first example in Table 1 is extracted from an autobiography. By converting from 1st to 3rd person PoV, the text could be used for a biography or for an encyclopedia, depending on the importance and the level of detail of the factual information conveyed. The second example is extracted from an educational text describing symptoms of a bipolar personality disorder. By converting the main character's PoV from 3rd to 2nd person, the text becomes more suitable for readers who are interested in interpreting the symptoms in the context of their own behavior, i.e. self-diagnosis. The third example shows a sample from the abstract section of a paper, where it is common for authors to use the 1st person PoV. By converting it to 3rd person, the transformed text becomes suitable for the related work section of another paper. Note that the direction of the PoV change can be reversed, transforming the second text into the first text in each of these example pairs.

The narrative examples from Table 2 have two major characters, or roles, E_1 and E_2, whose thoughts and behaviors are originally described using the second person point of view for E_1 and the third person point of view for E_2. The two roles are in a particular relationship, e.g. client vs. lawyer, patient vs. doctor, or child vs. toxic mother. In each example pair the narrative perspectives are switched such that role E_2 is assigned the second person PoV, while role E_1 receives the third person PoV, effectively changing the target audience of the text. Thus, the first text, which was originally intended to help clients recognize trustworthy lawyers, becomes more suitable for lawyers interested in earning the trust of their clients.

The adaptation of existing stories to satisfy the requirements of a new medium provides an additional application area. A large number of games are played from a second-person viewpoint (Harrigan and Wardrip-Fruin, 2007). However, many of these games are based on previously written stories where the protagonist uses the third-person PoV. Changing the narrative perspective could assist in the adaptation of the existing story in order to generate the game's introduction narrative or script.

In light of the above, in this paper we propose the task of *changing the narrative perspectives* of a text. We believe that developing a system that automatically changes the points of view would enable new types of narrative generation tasks. For example, a

E_1 in 2nd, E_2 in 3rd $\longleftrightarrow$ E_1 in 3rd, E_2 in 2nd
◆ Most cases aren't slam-dunks, and it is important that [*your lawyer*]$_2$ doesn't make promises regarding the outcome of [*your*]$_1$ case and should not be overconfident no matter how seasoned [*he or she*]$_2$ is. (Possinger, 2018).   ◆ Most cases aren't slam-dunks, and it is important that [*you*]$_2$ don't make promises regarding the outcome of [*your client's*]$_1$ case and should not be overconfident no matter how seasoned [*you*]$_1$ are.
◆ Communication begins from the moment [*you*]$_1$ first meet [*your doctor*]$_2$. Does [*he or she*]$_2$ greet [*you*]$_1$ warmly? (Center, 2018)   ◆ Communication begins from the moment [*your patient*]$_1$ first meets [*you*]$_2$. Do [*you*]$_2$ greet [*her*]$_1$ warmly?
◆ [*The toxic mother*]$_2$ is also likely to spot the speck in an otherwise perfect offering, and [*her*]$_2$ perfectionism will cause [*you*]$_1$ to never quite to feel good enough, no matter what [*you*]$_1$ do. (Platform, 2018)   ◆ [*You*]$_2$ are also likely to spot the speck in an otherwise perfect offering, and [*your*]$_2$ perfectionism will cause [*your child*]$_1$ to never quite to feel good enough, no matter what [*she*]$_1$ does.

Table 2: Changing from 3rd to 2nd person PoV in texts already containing a 2nd person PoV. Verbs whose conjugation is changed due to number and person agreement are shown with dotted underlines.

literature review generation system would change the narrative perspective from 1st to 3rd person in the abstract and conclusion sections that are extracted automatically from papers on a particular topic, possibly in tandem with a summarization or paraphrasing step. Another potential application would be the automatic generation or updating of Wikipedia articles using information extracted from biographies, where the 1st person PoV would need to be changed to the more neutral 3rd to make it suitable for an encyclopedic text. Although excluded from the task of *narrative* perspective change considered in this paper, the automatic processing of dialogues presents additional opportunities for a PoV change system. In one application, the system would be tasked with translating dialogues, where the 1st person PoV is pervasive, into an indirect speech version requiring a 3rd person PoV. This could be useful, for example, in generating meeting

6

notes or summaries of meetings where two or more people interact directly through dialogue.

1.2. Outline and main contributions

In the next section, we position the new task of changing the narrative perspective in the wider context of related work on text generation and narrative processing (Section 2). The following sections in the paper are then focused on its main contributions:

1. Characterizing the new task of changing the narrative perspectives of a text (Section 3).

2. Designing a neural architecture for mention selection (Section 5) that lies at the core of an end-to-end system for deictic to anaphoric PoV change, where the text is first pre-processed for mention detection, coreference resolution, and verb conjugation (Section 4).

3. Showing how coreference annotations can be used to create training examples for the mention selection task (Section 6.1).

4. Introducing a new benchmark dataset comprised of a diverse set of types of narratives that are manually annotated with deictic to anaphoric PoV changes (Section 6.2).

5. Evaluations showing the proposed architecture outperforms decision tree-based approaches and other baselines, by virtue of generating mentions that lead to text that is more fluent and less ambiguous (Section 7).

The paper ends with ideas for future work and concluding remarks (Section 8).

2. Related Work

Wiebe (1994) addressed the problem of tracking the psychological point of view in third-person fictional narrative text, which is about recognizing sentences that convey the thoughts, perceptions, and inner states of characters and mapping these psychological perspectives to the corresponding character in the story. Mani (2013) provided a

7

systematic overview of computational methods applied to the processing of core narratological concepts, such as narrative perspective, narrative threads, and narrative time. An annotation scheme, NarrativeML, was developed to facilitate the annotation of these concepts. More recently, Eisenberg and Finlayson (2016) trained SVM models to distinguish between first-person and third-person point of view in a narrative, as well as for determining whether the narrator is involved in the story, i.e. diegesis.

Most related to the PoV change is the task of generating referential expressions in context (GREC) (Belz et al., 2010). In one version of this task, Wikipedia articles provide the context, referential expressions to the main subject of the article are erased, and lists of possible referring expressions are provided. The task then is to select referring expressions to insert into the gaps such that the resulting text is fluent and coherent. In all GREC tasks, the PoV of the focus entity is always 3rd person, and the context, including mentions of other entities, does not have to change. The task of changing of narrative perspective, however, is more complicated in that the change from a deictic PoV (1st or 2nd) to an anaphoric PoV (3rd) may require changing the context. As explained in Section 3.1, deictic pronouns are agnostic of centering constraints, therefore changing them to anaphoric pronouns may lead to referential ambiguity, unless mentions of other entities are changed too. NeuralREG (Ferreira et al., 2018) is another relevant system, where the task is to generate referential expression in a fixed context, a delexicalized version of the WebNLG corpus (Gardent et al., 2017). In most instances, the context is composed of one or two sentences that describe in natural language the knowledge encoded in a set of RDF triples extracted from DBPedia. In contrast, the contexts in the PoV task are much longer and need to change to accommodate replacing deictic pronouns with anaphoric pronouns. Furthermore, NeuralREG uses a seq2seq decoder to generate referential expressions for entities that appear in both the training and test data. In contrast, by the definition of the PoV change task, the approach presented in this paper works with new entities at test time.

Also related to the PoV Change task is style transfer, where a text is transformed automatically from one particular style into another. The transferred style can refer to: formality (Heylighen and Marc Dewaele, 1999; Rao and Tetreault, 2018), modern English to Shakespeare English (Xu et al., 2012; Jhamtani et al., 2017), sentiment (Fu

8

et al., 2018; Shen et al., 2017), or gender and political slant (Prabhumoye et al., 2018).

Approaches to the PoV change task may benefit from techniques that are used in the general task of entity generation. Ji et al. (2017) presented a new type of language model, EntityNLM, which can model an arbitrary number of entities in context while generating each entity mention at an arbitrary length. Clark et al. (2018) introduced an approach to neural text generation that explicitly represented entities mentioned in the text, which was shown to benefit mention generation.

Overall, the new task of changing narrative perspectives can be seen as part of a larger interdisciplinary area concerned with computational approaches to storytelling and narrative processing. A number of diverse tasks being addressed in this area and related fields have been explored in the Storytelling Workshop series (Mitchell et al., 2018), as well as within this journal in its Special Issue on Narrative Extraction from Texts (Jorge et al., 2019). Finally, very recent developments in the field are to be presented in the upcoming Fourth International Workshop on Narrative Extraction from Texts (Text2Story) (Campos et al., 2021).

3. Task Definition and Important Aspects

We consider as input a text document in a prototypical PoV setting where at most one discourse entity has the 1st person PoV, at most one entity has the 2nd person PoV, and any number of entities have the 3rd person PoV[1]. Given a focus entity E that has PoV p, the task is to change the text such that E is assigned a different PoV q. Changing the PoV of the focus entity may also require changing mentions of other entities, as needed to eliminate ambiguity or to preserve the naturalness of the original text. We will use the term *confounding entities* to refer to entities other than the focus entity that are affected by the PoV change. In the first example shown in Table 1, the focus entity is "Akira Kurosawa". If the pronoun "him" were left unchanged, its referent would be ambiguous between *Akira* and the confounding entity *his brother*. The approach

---

[1]When counting the number of entities with a 1st or 2nd person PoV, we do not consider entity mentions within dialogues or direct speech, which are normally unaffected by the PoV change (as shown in some of the examples in this paper).

that will be described in Section 4 however has the capability to convert "him" to "her brother", eliminating part of the ambiguity.

Overall, the task of changing the PoV of a focus entity can be subcategorized into one of the following 3 categories:

- **Deictic to Deictic (D2D)**: PoV change from 1st to 2nd or from 2nd to 1st.

- **Anaphoric to Deictic (A2D)**: PoV change from 3rd to 1st or from 3rd to 2nd.

- **Deictic to Anaphoric (D2A)**: PoV change from 1st to 3rd or from 2nd to 3rd.

In this paper, we focus on the deictic to anaphoric D2A case for the reason that it is often also a required step in the A2D and D2D transformations. As will be shown in the examples below, A2D (and similarly D2D) changes can be done in three steps:

1. Substitution of a mention string with a deictic pronoun suitable for its case (e.g accusative) and number (e.g. singular);

2. Potential change in conjugation of a verb modifying the original mention string;

3. Deictic to anaphoric changes for confounding entities.

For example, given the coreference chains, changing an entity's PoV to the 2nd person starts by replacing all mentions in its chain with the appropriate 2nd person pronouns (step 1). However, the discourse may have already referred to another (confounding) entity using the 2nd person, in which case that entity's PoV needs to be changed to 3rd person to avoid ambiguity (step 3). In the examples from Table 2, in order to change the entity E_2 to use the 2nd person PoV, entity E_1, which was already using the 2nd person PoV, has to be changed to 3rd person PoV. Overall, changing the PoV may require, directly or indirectly, a D2A change from the 2nd person to the 3rd person. The same reasoning applies when the target PoV is 1st person. In the examples below, the A2D task is to change the entity *Cliff* from a 3rd to a 1st person PoV. In the first example, the substitution is straightforward: all the corresponding entity mentions, which are in the nominative case, are changed to the pronoun *I*:

- [*Cliff*]$_1$ → [**I**]$_1$ *laughed.* [*He*]$_1$ → [**I**]$_1$ *said, "I'm thinking of a cartoon I saw."*

10

In the second example, there is another (confounding) entity *Bill* that is already using the 1st person PoV. This will need to be changed to use a different PoV (by default 3rd person), to allow *Cliff* to take over the 1st person PoV. This example also shows the need to change the conjugation of a verb to agree in person with the new PoV:

- $[I]_2 \rightarrow$ [**Bill**]$_2$ *asked something, then* $[he]_1 \rightarrow$ [**I**]$_1$ *said something,* $[I]_2 \rightarrow$ [**Bill**]$_2$ *asked some more and* $[he]_1 \rightarrow$ [**I**]$_1$ *explained. The details don't matter, but as* $[I]_2 \rightarrow$ [**Bill**]$_2$ *said,* $[he]_1 \rightarrow$ [**I**]$_1$ *is* $\rightarrow$ **am** *the mathematician of the combination.*

The third example shows the need to choose a pronoun replacement appropriate for the accusative case:

- $[I]_2 \rightarrow$ [**He**]$_2$ *had been on the phone talking to* $[Cliff]_1 \rightarrow$ [**me**]$_1$ *in the lab.*

Note that the agreement in case and number (step 1) and the potential change in verb conjugation (step 2) required for A2D and D2D transformations are also necessary for a successful implementation of the D2A transformation. Therefore, given the need of the D2A transformation (as step 3) in A2D and D2D, we believe that successfully addressing D2A will also provide the solution for a significant number, if not the majority, of A2D and D2D transformations. This is not to say, however, that there are no situations in which A2D or D2D would require a more specialized approach: consider the example below, where the change from 3rd to 1st person PoV results in the somewhat awkward phrase *"a German in my mid sixties"*:

- [*This fellow*]$_1 \rightarrow$ [**I**]$_1$ *was the son of a Nazi – a German in* $[his]_1 \rightarrow$ [**my**]$_1$ *mid sixties whose father had served in the SS during World War II.*

The example above further illustrates the complexity of the task of changing the narrative perspective. Ideally, a complete solution would have the capability to also transform the text between the mentions whenever that can preserve the naturalness and fluency of the original text. As a first step in solving the PoV change task, in this paper we consider a formulation where the change in narrative perspective is done by changing only mentions of focus entities and their confounding entities, leaving the more general transformation of text as a direction for future work.

The deictic to anaphoric PoV change that is the focus of this paper is a non-trivial task that requires selecting a third person pronoun or noun phrase to replace mentions of the focus entity E such that the resulting text satisfies two major criteria:

○ **Manageable ambiguity**: The selected strings lead to non-ambiguous references, e.g. when a pronoun is generated, the reader can easily and correctly infer the entity it refers to.

○ **Natural text**: The text sounds natural, which generally means that it is grammatical and not repetitive in terms of the words used for mentions of the same entity.

In Section 7.2 we provide examples illustrating the various levels of referential ambiguity and naturalness used for the human evaluations. These examples were included in the annotation guidelines in order to help the human annotators reliably assign referential ambiguity and naturalness scores.

3.1. Interactions with Other Aspects of Language

The task of changing the PoV can be further complicated by *quotations*, *dialogues*, *generic pronouns*, and the use of *narrator's perspective*.

The text inside quotations and dialogues is usually kept unchanged when converting the PoV, however identifying dialogues and quotations can be difficult Pareti et al. (2013), especially when no formatting, e.g. quotation marks, is available. Likewise, pronouns used in a generic sense are kept unchanged when converting the PoV, however identifying generic expressions is still a very difficult NLP task – Friedrich and Pinkal (2015), for instance, achieve a state-of-the-art of less than 60% F1 score. The PoV also needs to stay unchanged for the narrator's perspective, as will be shown in the *Mother Night* Vonnegut (1999) example from Section 6.2. There, it is clear that the narration uses the voice of the author, and thus the PoV needs to stay in the 1st person.

Smith (2003) catalogued five discourse modes: *Narrative, Description, Report, Information* and *Argument*. The events in Narratives are temporally related according to narrative time, whereas in Descriptions the text progression is spatial rather than temporal. In Reports, the time of the situations expressed is evaluated against speech

time, whereas the Information and Argument modes are atemporal. The dataset created for the task of PoV change is mainly based on texts in the Narrative mode, where the alternation between different timelines may introduce additional difficulties – such an example will be shown in Section 6.2, where the narrative alternates between the past (the time of the events in the story, where the PoV changes) and the present (the narrator's time, where the author's voice is used extensively and therefore the original PoV needs to be preserved).

In this paper we focus on changing the PoV in English. While many aspects of the proposed approach can transfer to other languages, changes may be required depending on properties of the actual language. For example, in pro-drop languages, pronouns in subject position are allowed to be dropped in certain contexts. Therefore, in addition to overt pronouns and noun phrases, the mention selection step would also need to consider zero pronouns. There is usually a flexible word order in languages that have extensive case systems and use a high degree of morphological marking to disambiguate the semantic roles of the arguments. In such languages, word order changes may be required when changing the PoV. For example, when replacing a first person pronoun with a proper name, the case and morphological marking as well as position may differ from the original. Furthermore, coming up with unambiguous references for PoV changes will be more difficult in languages that do not make gender distinctions.

Similar to the task of generating referential expression in context (GREC) (Belz et al., 2010), the task of changing the PoV requires selecting the form of referential expression: pronoun vs. description vs. proper name. The Centering Theory (Grosz et al., 1995) relates focus of attention, choice of referring expression, and perceived coherence of utterances within a discourse segment. Accordingly, discourse entities that are *salient* in the current context are more likely to be realized in text using shorter expressions, such as pronouns. A similar conclusion is reached using other psycholinguistics models of reference, such as the Uniform Information Density (UID) hypothesis according to which speakers tend to convey information at a uniform rate (Jaeger, 2010; Jaeger and Buz, 2016). Applied to the choice of referring expressions, UID entails that a highly *predictable* referent should be encoded using a short code, e.g. a pronoun,

while an unpredictable referent should be encoded using a longer form[2]. However, unlike the GREC task where the PoV is always 3rd person in a context that is given and does not need to change, in the PoV change task the context might have to change because of the switch between deictic pronouns (1st or 2nd) and anaphoric pronouns (3rd). In the context of centering theory, Poesio et al. (2004) found that 2nd person pronouns behave differently from 3rd person pronouns. One rule in centering says that if any of the forward-looking centers (CF) are realized as a pronoun in the current sentence, then its backward-looking center (CB) should be pronominalized too. However, the corpus-driven study of Poesio et al. (2004) found that the rule does not apply to 2nd person pronouns, confirming Walker (1993) who had suggested that they were not covered by centering. This is important for the task of PoV change, because a context that was structured for a 2nd person PoV (which is agnostic of centering rules) might have to be altered when changed to a 3rd person PoV to satisfy the centering rules and constraints, which specify conditions for the coherence of the resulting text. For example, in fixed-word-order languages such as English, syntactic position determines the partial ordering of the CF list, with subject ranked higher than objects, which in turn are ranked higher than other positions. When a new 3rd person pronoun is introduced in a sentence, this might break the centering rule above, because syntactic positions positions are the same, and hence the ordering of the CF list is unchanged. Allowing other pronouns in the sentence to be replaced can alleviate this, as shown in the first example in Table 1, demonstrating the importance of developing a model that can change not only the focus entity mentions, but also realizations of other entities. In (relatively) free-word-order languages such as German, functional rather than syntactic criteria are important for the ranking of CF lists (Strube and Hahn, 1999), based on the distinction between hearer-old and hearer-new discourse entities (Prince, 1981). This seems to indicate that a PoV system designed for English might need to be adapted to work for free-word-order languages.

---

[2]For a more in-depth discussion of UID and factors that affect prediction, the reader is referred to (Modi et al., 2017)

4. The PoV Change System Pipeline

To change the narrative perspective from $1^{st}/2^{nd}$ to 3rd PoV, the raw text is processed using the pipeline of 4 steps shown below. As a running example, we will use the text below, where the focus entity (to be indexed with 2, and later named *Nick*) has the 1st person PoV (as expressed through the pronoun *I*) and the task is to convert it to a 3rd person PoV:

> *"Phil returns to Boston, to his job. I drive to the city every other week, to work a night or two at Pine Street, to see Emily."* (Flynn, 2005)

Step 1: Identify coreference chains corresponding to the focus and confounding entities.

1.1) Perform coreference resolution (Lee et al., 2017) and named entity recognition (Honnibal and Montani, 2017) to extract all coreference chains referring to people.

> $[Phil]_1$ *returns to Boston, to* $[his]_1$ *job.* $[I]_2$ *drive to the city every other week, to work a night or two at Pine Street, to see* $[Emily]_3$.

1.2) Identify the coreference chain corresponding to the focus entity, based on the original PoV or the name of the focus entity.

> $[Phil]_1$ *returns to Boston, to* $[his]_1$ *job.* $[\mathbf{I}]_2$ *drive to the city every other week, to work a night or two at Pine Street, to see* $[Emily]_3$.

1.3) Coreference chains that contain singular third person pronouns that match the gender of the focus entity will be considered confounding entities. We assume the name and the gender of the focus entity are given as input. In general, the gender cannot be determined from the text only, as deictic pronouns can refer to any gender.

> $[\mathbf{Phil}]_1$ *returns to Boston, to* $[\mathbf{his}]_1$ *job.* $[I]_2$ *drive to the city every other week, to work a night or two at Pine Street, to see* $[Emily]_3$.

In sub-step 1.3, an additional set of confounding entities will be created from coreference chains that contain plural pronouns that have the original PoV of the focus entity, e.g. if the original PoV is 1st person, then a chain containing "we" will be considered as a confounding entity. Mentions inside quotations and dialogues are ignored, as well

as pronouns that do not belong to any coreference chain (e.g. generic pronouns). First person pronouns that refer to the narrator are ignored as well by searching for patterns indicating a narrator's perspective, such as "I promise you", "I guarantee you", "I guess". A list of performative verbs (Underwood, 2009) are used to generate these patterns.

Step 2: Change conjugation of verbs with focus entity mentions as subjects.

2.1) Identify verbs whose subject is a focus entity mention and change their conjugation if necessary such that they agree in person with the new PoV, as shown underlined in Table 1 and in bold in the example below.

[*Phil*]$_1$ *returns to Boston, to* [*his*]$_1$ *job.* [*I*]$_2$ **drive** → **drives** *to the city every other week, to work a night or two at Pine Street, to see* [*Emily*]$_3$.

To identify subject relations, we use a set of rules that leverage syntactic dependencies extracted by the dependency parser of Dozat and Manning (2017).

2.2) Verb conjugation is changed using the dictionary from (Manning et al., 2014), or a set of simple rules for out of vocabulary verbs, such as add 'es' to a verb ending in 's','sh','ch','x','z',or 'o'.

Step 3: Create a set $S(E)$ of candidate mention strings for each focus or confounding entity E.

3.1) For a focus entity E, we first add to $S(E)$ their full name, first and last name, whenever known. We then add pronouns that match their gender, i.e. either {*she, her, herself*} or {*he, him, his, himself*}. We further leverage the syntactic dependencies extracted with the parser used in step 2 to identify predicate nominals and appositives that involve the focus entity and add them to the candidate set, e.g. "the doctor" from "I was a doctor". Finally, we use a dictionary of relational nouns (Newell and Cheung, 2018) to identify all possessive constructions where the focus entity modifies a relational noun, e.g. "my sister". Depending on the gender of the focus entity, we then add to its $S(E)$ the converse relational phrase, e.g. either "her brother" or "her sister". Furthermore, if the confounding entity "my sister" is coreferential with a name, e.g. "Mandy", we add the corresponding possessive construction too, e.g. "Mandy's brother" or "Mandy's sister".

3.2) For a confounding entity E that is 3rd person singular, its candidate set $S(E)$ will contain the strings of all the mentions in its coreference chain. Confounding entities that have plural 1st or 2nd person PoV need to be changed to plural 3rd person PoV, as such their $S(E)$ will contain all plural third person pronouns. Additionally, coordinated phrases that contain names will be transformed appropriately and added to the candidate set, e.g. an original occurrence of "Mandy and me" that corefers with "we" leads to "Mandy and him" or "Mandy and her" being added to $S(E)$, depending on the gender of the focus entity.

Setting E_1 to be the confounding entity *Phil* and E_2 to be the focus entity *Nick*, using the procedure above their candidate sets will be created as: $S(E_1) = \{$*Phil*, *his*$\}$ and $S(E_2) = \{$*Nick Flynn, Nick, Flynn, Nick Flynn's, Nick's, Flynn's, he, him, his, himself, he himself, his son, her son, his brother, his grandson*$\}$.

Step 4: Going from left to right, for each focus or confounding entity mention, select an appropriate string from $S(E)$, using the mention selection approach described in Section 5.

4.1) For each mention, we also use the original mention string to narrow down the choices in $S(E)$. For example, if the original mention string is a reflexive or possessive pronoun (e.g. "my" in "my sister"), then the choices of potential mention strings are constrained to reflexive pronouns or possessive pronouns and nouns, respectively (e.g. "Akira's" in "Akira's sister").

Using the same example, the final result will be:

[*Phil*]$_1$ *returns to Boston, to* [*his*]$_1$ *job.* [*I*]$_2$ → [**Nick**]$_2$ *drive* → *drives to the city every other week, to work a night or two at Pine Street, to see* [*Emily*]$_3$.

5. A Neural Mention Selection Model

Given a mention position t in the text for an entity E, the mention selection task is to select from the set $S(E)$ of possible strings for that entity the best string (noun phrase or pronoun) to use as a mention M_t in that context. Using the same example as in the previous section, where the focus entity E is the one indexed with 2, and assuming that

it was already mentioned 5 times before the sentences in this example, then the current mention position would be 6, as shown below:

[*Phil*]$_1$ *returns to Boston, to* [*his*]$_1$ *job.* [M_6]$_2$ *drives to the city every other week, to work a night or two at Pine Street, to see* [*Emily*]$_3$.

Its corresponding set of candidate string, narrowed down in sub-step 4.1 above, will be $S(E_2) = \{$*Nick Flynn, Nick, Flynn, he, he himself, his son, her son, his brother, his grandson*$\}$.

We train a scoring function $score(s|C)$ to capture how appropriate a string $s \in S(E)$ is to be used as a mention of an entity E in a context C. Given $s_i, s_j \in S(E)$, we use $\langle s_i > s_j | C \rangle$ to denote that s_i is more appropriate as a mention string than s_j in the context C. Correspondingly, the ranking system is expected to rank s_i higher than s_j, by computing a $score(s_i|C) > score(s_j|C)$. In the example, this means that the system is expected to compute $score(Nick|C) > score(he|C)$ for mention M_6.

In our experiments, we use the margin-based ranking loss shown below, where γ is a tunable margin hyper-parameter:

$$J(S(E)|C) = \sum_{s_i > s_j | C} max\{0, \gamma - score(s_i|C) + score(s_j|C)\} \tag{1}$$

At training time, we use the observed mention string $\hat{s} \in S(E)$ to create ranking pairs $\langle \hat{s} > s | C \rangle$, for all $s \in S(E), s \neq \hat{s}$, as training examples.

Figure 1 shows the neural architecture trained to compute the ranking $score(s_j|C)$. Given a mention position t in the text, we split the context into a left context C_{left} and a right context C_{right}, i.e. $C = [C_{left}, C_{right}]$, which are processed sequentially using two LSTM networks (Hochreiter and Schmidhuber, 1997):

- **Token-level**: $LSTM_{left}^T$ processes the N tokens in the left context C_{left} from left to right, followed by the tokens in the candidate string s_j. $LSTM_{right}^T$ processes the N tokens in the right context C_{right} from right to left, stopping right after the string s_j. The final states from the two LSTMs are concatenated to produce a token-level representation $h^{(T)}$. In the example above, if $N = 5$ and s_j="Nick", then $LSTM_{left}^T$ will process $\langle$",", "to", "his", "job", ".", "Nick"$\rangle$ in this order, whereas $LSTM_{right}^T$ will process $\langle$"every", "city", "the", "to", "drives"$\rangle$ in this order.
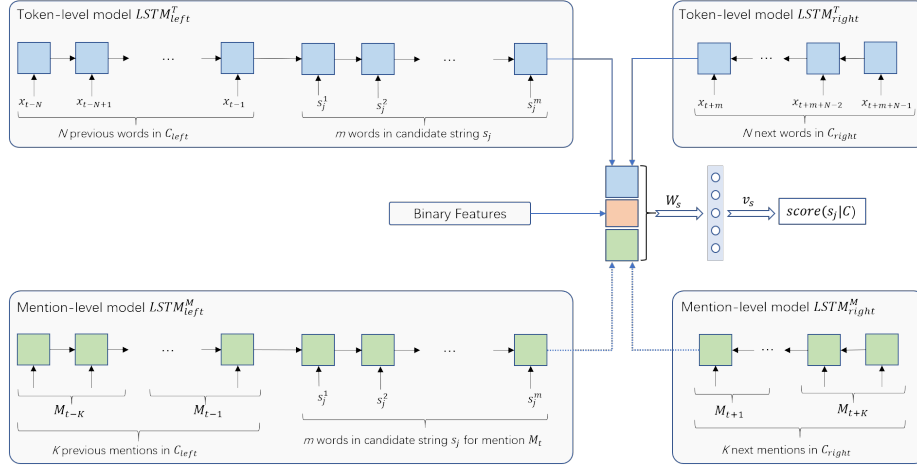
Figure 1: Computation of $score(s_j)$ using token-level and mention-level LSTMs for left and right contexts.

○ **Mention-level**: $LSTM_{left}^M$ processes the tokens from the K mentions in the left context C_{left}, from left to right, followed by the tokens in the candidate string s_j. $LSTM_{right}^M$ processes the tokens from the K mentions in the right context C_{right}, from right to left. The final states from the two LSTMs are concatenated to produce a mention-level representation $h^{(M)}$. In the example above, if $K = 2$ and s_j="Nick", then $LSTM_{left}^M$ will process ⟨"Phil", "<sep>", "his", "<sep>", "Nick"⟩ in this order, whereas $LSTM_{right}^M$ will process ⟨"<pad>", "<sep>", "Emily"⟩ in this order.

The use of LSTMs to process the context words and mentions is motivated by the fact that sequential information is essential for mention selection. For example, it is important to know whether a pronoun was used most recently for a different entity in the current sentence vs. two sentences before, because a pronoun should have a higher likelihood to be selected if the later were true, as it is less likely to create ambiguity while maintaining naturalness. Theoretically, when augmented with the binary features described below and when using a large value for the context size N, the token-level LSTM should be able to capture the relevant mention-level information. However, in practice, LSTMs can still "forget" relevant information from across large spans of text. Running the second LSTM only over the mention sequence is intended to alleviate this.

19

The ablation results shown in Table 6 later in Section 7 further validate this design choice.

Because the mention-level LSTM processes tokens from mentions, and because mentions may have a variable number of tokens, a special token $\langle sep \rangle$ is added before each mention to mark mention boundaries. Since at test time the models will be run from left to right, this means that the strings for the relevant mentions at the right of the current mention are unknown. As such, they are replaced with a special token $\langle unk \rangle$ that receives its own embedding. For each token that is processed by the token-level and mention-level LSTM, we use as input their contextualized embeddings provided by BERT Devlin et al. (2019). Additionally, for the mention-level LSTMs, the BERT embedding for each token is augmented with a vector of binary features ϕ_t^b computed as follows:

1. Whether the mention belongs to the focus entity.

2. A one-hot vector for the distance d in tokens between the current mention and the previous mention. A binary feature is created for each of the following intervals: $d \leq 5, 5 < d \leq 10, 10 < d \leq 15, 15 < d \leq 20, 20 < d \leq 25$, and $d > 25$.

These two features aim to help the system learn to avoid ambiguity and repetition. For example, if the mention M_{t-1} right before M_t is expressed as a pronoun, it does not refer to the focus entity, and the distance between the two mentions is small, then the system should learn to use a noun phrase for M_t in order to avoid referential ambiguity.

In addition to the states computed by the LSTMs, we also use a vector ϕ^b of binary features, as follows:

1. Whether M_t is the first or second mention in the coreference chain for E.

2. A one-hot vector for the mention length, i.e. the number of words contained in s_j. A separate binary feature is created for each length in $[1, 5]$ and one binary feature for lengths greater than 5.

3. Whether the candidate string s_j for mention M_t has already been used for a previous mention in the same coreference chain.

4. Whether s_j has been used for the previous mention M_{t-1} in the coreference chain for entity E.

5. Whether M_t is the subject or object of a verb.

The first two mentions in a coreference chain are usually longer noun phrases. Features 1 and 2 are expected to help the system learn to choose longer noun phrases for the first two mentions. Through features 3 and 4, the system is expected to learn to avoid repetition by using different strings. The information of whether the mention is a subject or object provided through feature 5 should help the system to choose grammatically correct mention strings, for instance avoid selecting the object pronoun "him" if the mention is in a subject position, as in "*Bill* told Jake to go to work". Note that in this example the original string *Bill* on its own cannot give any information as to whether the original mention string was in a subject or object position. Note that with the exception of feature 2, all binary features rely on coreference or syntactic dependency information, which is unlikely to be learned implicitly by the LSTM on such a small dataset, hence the utility of encoding these features explicitly.

The states from the token-level and mention-level LSTMs are concatenated with the binary features to create an overall feature vector $\phi(s_j, C) = [h^{(T)}|h^{(M)}|\phi^b]$. Following the ranking approach of Collobert et al. (2011), this feature vector is then fed into a fully connected network with one hidden layer (W_s, b_s) using *tanh* as non-linearity, followed by a linear output transformation (v_s) into a final ranking score as follows:

$$score(s_j|C) = v_s^T \tanh(W_s\phi(s_j, C) + b_s) \qquad (2)$$

6. Datasets for PoV Change in English

Training deep learning models, such as the neural mention selection architecture that we propose for this task, requires a large number of training examples. However, manually annotating the corresponding large number of documents with PoV changes is prohibitive. Instead, we propose to use existing coreference resolution corpora to create training examples for the mention selection part of the PoV change task, as described in Section 6.1 below. Separately, we also create a benchmark dataset containing manual annotations of PoV changes, to be used only during testing, as described in Section 6.2.

6.1. Using Coreference Chains for Training

The CoNLL-2012 corpus (Pradhan et al., 2012) is a widely used dataset for the training and evaluation of machine learning approaches to coreference resolution. The dataset adds a coreference annotation layer to the OntoNotes corpus Hovy et al. (2006); Weischedel et al. (2011). The coreference layer annotates general anaphoric relations covering a rich set of entities and events that are not limited to noun phrases or a limited set of entity types. In this paper, we use the English portion of the dataset, which is comprised of documents from various genres of text, ranging from news, to conversational telephone speech, weblogs, Usenet newsgroups, broadcast, and talk shows. For each document in the corpus, we select only the coreference chains that correspond to person entities and have the 3rd person PoV. Each of these coreference chains can be used as the focus entity, which is assumed to have been transformed from 1st to 3rd person PoV, while the remaining chains that agree in number and gender are used as confounding entities. Since input documents for the PoV change task should contain at most 1 entity with 1st or 2nd person PoV, as per the task definition ins Section 3, this requires removing from the CoNLL corpus all documents that contain deictic PoVs. We use the official partition of the CoNLL-2012 corpus into training, validation, and testing. Upon filtering out documents with deictic PoVs, the statistics of the remaining documents in terms of person entities and their mentions are shown in the top 3 rows of Table 3. Overall, about half of the original CoNLL dataset is used in our experiments. Given a coreference chain $E = \{m_1, m_2, ..., m_T\}$, we create $S(E)$ by collecting the unique strings that are used across all the mentions in the chain. Then the mention selection part of the PoV task can be reduced to selecting, for each position t, the actual string in $S(E)$ that was used to express mention m_t.

6.2. The PoV Change Dataset

To enable the evaluation of various approaches to PoV change task, we created a separate corpus of documents in which the PoV change was done manually for the main character in each document. The major criteria for selecting documents to include in this benchmark dataset were as follow:

22

Dataset	Entities	Mentions	Men / Ent	Docs	Words
CoNLL-train	1,952	11,727	6.0	1,143	603,885
CoNLL-dev	267	1,446	5.4	154	80,181
CoNLL-test	253	1,496	5.9	124	73,582
PoV-change	300	8,682	28.9	21	84,312

Table 3: The number of person entities and mentions, average mentions per entity (Men / Ent). as well as the number of documents and words in each dataset.

1. The text should contain a first- or second-person narrative with a sufficiently large number of mentions (at least 10), such that a mention selection component would have to make non-trivial decisions. This typically means generating names, nouns, and pronouns that refer to the corresponding entity unambiguously, while maintaining the naturalness of the text.

   - Because dialogues an direct speech are left unchanged (Section 3.1), 1st and 2nd person pronouns that appear in such contexts are not counted.

2. To enable the evaluation of a system's capability to avoid referential ambiguity, the text should also contain one or more confounding entities, with a non-trivial total number of mentions (at least 5). This means that the narrative should have one or more characters of the same gender as the focus character, which for the first-person PoV would be the narrator.

3. The corpus should contain texts written in different styles, in order to capture a diverse set of referential devices and situations.

Upon a thorough exploration of the texts in the original (full) CoNLL corpus, we found that only 3 texts satisfied these criteria: one from the training portion (a2e_0015), one from development (c2e_0040), and one from test (c2e_0029). In the broadcast news part of the corpus, for example, first and second person pronouns appear solely in dialogues in direct speech. Telephone conversations, talk shows, and broadcast conversations consist solely of dialogues, whereas the only situations in which first and second person points of view appeared in news and magazine texts were in direct speech. The only 3 texts that satisfied the corpus inclusion criteria were all found in the Usenet

newsgroups part of CoNLL. In Usenet too the 1st and 2nd person pronouns are predominantly within dialogues and direct speech, although the dialogues are often implicit: a newsgroup post either responds to the author of an earlier post or addresses all readers of the newsgroup using 2nd person pronouns. These types of narratives also often contain numerous *imperative statements*, e.g. "Excuse me if I am unable to answer", "Do not misunderstand me" (c2e_0024), "I wish that you were more logical, my dear brother" (a2e_0011), "If you think i am exaggerating things to scare people, then shut up", "Think of whether you have actual power to" (c2e_0034), "Please guard your moral bottom line"(c2e_0033), or *rhetorical questions*, e.g. "What do you need?", "Do you get excited?", "Are you worried?", "Are you determined?" (c2e_0012). For both imperative statements and rhetorical questions it is unclear whether a PoV change would make sense.

Since the training part is used to train the mention selection methods, from the CoNLL corpus we manually annotated only the document c2e_0040 from development and c2e_0040 from test that satisfied the corpus inclusion criteria (making sure to eliminate c2e_0040 from development during hyper-parameter tuning). We completed the PoV change corpus by manually annotating an additional set of 19 texts, which were taken from the sources listed in Table 4. In 18 texts the PoV was changed from 1st to 3rd person, whereas in the text from Evans (2015) the PoV was changed from 2nd to 3rd person. The inclusion of only one text with a 2nd person PoV is a consequence of the fact that in narratives the second person point of view is much less common, compared to first person. The manual annotation was done by a graduate in Linguistics.

The texts were selected to provide a mix of fiction and non-fiction, as well as a combination of short stories, web articles, and book excerpts, and not be dominated by dialogue. For instance, *Sweetness* Morrison (2015) is an emotional confession of a mulatto mother about how she treated her black daughter. She often addresses the reader directly, which makes it hard to change the PoV without modifying more of the text. *Mother Night* Vonnegut (1999) is a novel written as a fictional memoir of a former Nazi propagandist named Howard Campbell. The narrative alternates between present and past and the narrator's voice is heavily present, which makes this a more complex text. For example:

Source	Genre / Type	Cf	M_{cf}	MC_e	P	R	F_1
Chu (2013)	Fiction / Short story	14	33.4	3.5	70.3	73.8	72.0
Chiang (1991)	Fiction / Short story	7	33	4.9	59.5	65.4	62.3
Herr (1977)	Non-fiction / Book excerpt	37	10.5	3.4	71.7	72.0	71.8
Flynn (2005)	Fiction / Book excerpt	34	13.5	3.3	73.8	76.5	75.1
Bryson (2010)	Non-fiction / Book excerpt	13	17.7	2.6	71.4	72.2	71.8
Pollan (2018)	Fiction / Book excerpt	17	12	2.8	83.9	81.7	82.8
Solnit (2013)	Non-fiction / Book excerpt	20	20.2	3	73.8	72.7	73.2
Samatar (2013)	Fiction / Short story	15	19.9	3.8	74.0	74.5	74.2
Vonnegut (1999)	Fiction / Book excerpt	18	11.1	3.7	70.3	67.0	68.6
Asimov (1953)	Fiction / Book excerpt	7	41.3	4.1	73.5	72.5	73.0
Koul (2016)	Non-fiction / Short story	9	16.2	3.1	73.5	74.4	73.9
Evans (2015)(∗)	Non-fiction / Web article	4	57.5	5.0	62.8	69.2	65.8
CoNLL-c2e-0029	Non-fiction / Web forum	6	9.7	3.6	74.2	72.0	73.1
CoNLL-c2e-0040	Non-fiction / Web forum	1	4.0	2.5	82.8	80.0	81.4
Ungerleider (2016)	Non-fiction / Web article	9	17.7	4.2	71.5	73.6	72.5
Morrison (2015)	Fiction / Short story	16	11.4	3.4	74.3	77.7	76.0
Biss (2009)	Non-fiction / Book excerpt	10	9.1	3.1	68.1	67.0	67.5
Miller (2016)	Fiction / Short story	9	6.3	2.9	81.9	80.5	81.2
Atwood (1998)	Fiction / Book excerpt	17	27.4	3.6	68.9	70.2	69.5
Key (2015)	Non-fiction / Short story	9	21.4	3.8	74.0	73.2	73.6
Sartre (1969)	Fiction / Book excerpt	7	16.7	4.1	70.0	68.3	69.1

Table 4: The sources used for the PoV change dataset. Cf = total number of confounding entities; M_{cf} = average number of mentions per confounding entity; MC_e = average number of mention candidates per entity; P, R, and F_1 are the precision, recall, and F-measure of the end-to-end PoV change system on each document.

- *When he → Kraft opened up, I → Campbell saw that he was a painter. There was an easel in the middle of his living room ...*

- *When I talk about Kraft, alias Potapov, I'm a lot more comfortable than when I talk about Wirtanen, who has left no more of a trail than an inchworm crossing a billiard table. Evidences of Kraft are everywhere.*

In the first paragraph above, the writer recounts Campbell's story, using the past tense. Then, in the second paragraph, he switches to the present tense, providing additional information about one of the characters and expressing his feelings regarding him. As this is clearly the voice of the author, it needs to stay in the 1st person.

To enable assessing the performance of mention selection, these articles were also manually annotated with coreference chains and verb conjugation changes. The PoV change dataset is used solely for evaluation purpose and was originally partitioned into validation and testing, as shown in Table 4 in the top 14 rows and bottom 7 rows, respectively. Eventually, as tuning on the development portion of the PoV dataset did not lead to better mention selection accuracy when compared with tuning on CoNLL development data, the PoV change results were computed over all 21 documents. The statistics of the full PoV dataset are shown in the last line of Table 3. While the number of texts may seem small, the number of words is comparable (slightly larger) than in CoNLL-test. This is because the PoV documents are much larger than those in CoNLL. On average, a PoV document has 4,015 words, whereas the CoNLL-test document is on average only 593 words long. Out of the 19 PoV documents that are not from CoNLL, the smallest document is (Biss, 2009) with 1,106 words, whereas the largest is (Chu, 2013) with 7,611 words. Furthermore, the number of person entities is larger in PoV than in CoNLL-test, whereas the total number of mentions in PoV is more than 3 times the number of mentions in CoNLL-test. Coupled with the fact that PoV documents are much longer, this results in an average number of mentions per person entity in PoV that is almost 5 times larger than in CoNLL-test. The longer coreference chains in the PoV dataset are important in that they enable evaluating the ability of PoV change systems to select mention strings that are not repetitive or ambiguous over larger spans of texts, and in the presence of multiple confounding entities.

Given an input text and a focus entity, there may be multiple solutions to the task of PoV change, each a with different trade-off between the manageable ambiguity and naturalness criteria. To get a sense of the potential differences, a fragment from the autobiography of Kurosawa (1983) was annotated independently by two annotators. There are 169 focus entity mentions in the annotated fragment. In 121 mention conversions both annotators used the same pronouns, whereas in 22 mention conversions both used the same noun phrases – including one situation where both used a noun plus pronoun combination, but one chose to use anaphora and the other cataphora. There are 26 cases where one chose to use pronouns and the other chose to use noun phrases, as in the example below:

{I} → {Akira} had been one year old at the time. The dimly lit place where {I} →
*{**he**} / {**the baby**} sat in a tub lodged between two boards was the room that served*
as both kitchen and bath in the house where {my} → {his} father was born.

In most cases the two annotators chose to use noun phrases during transition between
past and present, then for clarity, and lastly for stylistic reasons. Overall, the inter-
annotator agreement was 84.6%.

7. Experimental Evaluation

The training portion of the CoNLL-2012 corpus is used to train two LSTM-based
models: a model that uses only the token-level LSTMs, and a model that uses both the
token-level and mention-level LSTMs. The size of the hidden state of LSTM cell is set
to 50, whereas the size of the hidden layer of the fully connected network that computes
the ranking score is set to 100. We use $N = 50$ words and $K = 10$ mentions for both
the left and right contexts. If there are less than N words or K mentions available, we
use a special token $\langle pad \rangle$ with its own embedding to pad the sequence.

The models are implemented in PyTorch. The learning rate, the margin γ, and the
dropout rates are tuned using grid search on the validation portion of CoNLL-2012,
which is also used for early stopping. The models are then tested separately on the
corresponding test portion of CoNLL and the PoV dataset. Because models that were
originally tuned on the PoV development portion obtained the same results as models
tuned on CoNLL-dev, we combined the development portion of PoV with the test por-
tion and report results on the entire set of 21 documents using models trained and tuned
solely on CoNLL. Optimization is performed with the Adam algorithm Kingma and Ba
(2015). Testing is done in auto-regressive mode, by doing mention selection from left
to right. The models were tuned and trained on a desktop computer with an Intel Core
i7-8700K CPU @3.70GHz with 16GB memory, and an NVIDIA GeForce GTX 1080
with 8GB memory. Training the tuned model on this platform took 16 hours and 50
minutes.

	Mention Selection		PoV End-to-End		
Model	CoNLL	PoV	P	R	F1
LSTM: Token + Mention	**74.1**	**75.2**	**71.6**	**73.0**	**72.3**
LSTM: Token	70.9	73.4	69.7	71.1	70.4
Decision Trees	52.7	57.2	49.1	50.0	49.5
Random Forests	57.7	59.8	54.7	55.8	55.2
Gradient Boosted Trees	64.8	62.9	57.2	58.3	57.7
Random Selection	33.9	33.6	21.6	22.1	21.8
Only Pronouns	28.6	42.5	37.1	37.8	37.4
Most Common String*	45.4	48.8	37.1	37.9	37.5

Table 5: Comparative model performance (%) in terms of accuracy for Mention Selection, Precision (P), Recall (R), and F1 score for End-to-End.

7.1. Automatic Evaluation

Models are evaluated using precision (P), recall (R), and F_1 score with respect to both the mention and verb conversions, using the manual annotations as ground truth. Thus, if the system changed a total of N words, and N_1 of them were correctly changed, then precision is calculated as N_1/N. If the ground truth specifies that N_2 words should have been changed, then recall is calculated as N_1/N_2. The F_1 measure is then calculated as the harmonic mean of precision and recall. For the mention selection evaluations, since the number of mentions to change is known, i.e. N is always the same as N_2, we report accuracy as the percentage of mentions that were changed to the ground truth string. Table 5 reports two types of test results: *End-to-End* refers to the performance of the end-to-end system, where mention detection, coreference resolution, verb conjugation, and generation of $S(E)$ are done automatically by the system; *Mention Selection* refers to the performance of the mention selection component, using the manual annotations for everything else. Since the CoNLL dataset does not have the verb conjugation annotation for the 1st person PoV, we report only the mention selection performance for it. Furthermore, the End-to-End performance of the full *LSTM: Token + Mention* model on each document in the PoV dataset is presented in the last 3 columns of Table 4.

28

We also present the system performance when mention selection is done using 2 simple baselines and 3 trained decision tree-based approaches. The *Random Selection* baseline chooses a string from $S(E)$ at random. The *Only Pronouns* baseline uses only pronouns, making sure that they satisfy gender, number, and case agreement. We also present the results of a *Most Common String* approach which uses only the most common string across all mentions of E. Note that this approach cannot be used in practice because it needs to know the true counts of strings from $S(E)$, which is unknown at test time. We also implemented mention selection using 3 decision tree-based approaches: *Decision Trees, Random Forests*, and *Gradient Boosting Trees*. A set of binary features are calculated for each of the 10 mentions before and after the current mention, as follows:

1. The distance d in words between the mention and the current mention. A separate binary feature is created for each of the following distance intervals: $d \in [0, 5]$, $d \in (5, 10]$, $d \in (10, 15]$, $d \in (15, 20]$, $d \in (20, 25]$, $d > 25$.
2. If the mention string is a pronoun, proper noun or common noun.
3. If the mention belongs to the same entity as the current mention.
4. If the mention belongs to the same sentence as the current mention.
5. If the mention has the same number and gender as the current mention.
6. If the mention string is the same as the string from $S(E)$ used for the current mention.
7. The mention length l in words. A separate binary feature is created for each value of $l \in [1, 5]$ and for $l > 5$.
8. If the mention is $\langle pad \rangle$.

A similar set of binary features are extracted for the 10 mentions that appear before the current mention and belong to the same entity as the current mention. A BERT score is calculated for each mention string from $S(E)$, and used as a feature to indicate how likely it is for that string to appear in the context for the current mention. Other features used for the current mention include its part of speech, its length, and whether it is the subject or object of a verb. Finally, we use the original mention string to narrow down the choices in $S(E)$, as described in Section 5. Note that the binary features used by the decision tree-based models subsume all the binary features used in the mention-level

29

and token-level LSTMs.

The decision tree-based models are trained as binary classifiers. For each $s \in S(E)$, if s is the correct string for the current mention, then the label is positive, otherwise negative. During testing, for a given mention position, the string s from $S(E)$ that obtains the highest probability to be positive is selected as the correct mention string. We used decision trees as a baseline as we were also interested in the performance of ML models that can ultimately be encoded as rules. Furthermore, ensembles of decision trees such as Random Forests and Gradient Boosted Trees are known to be very effective ML models.

The results in Table 5 demonstrate the importance of modeling the entities separately through the mention-level LSTMs, as the full Token + Mention model outperforms the Token model on both datasets. The mention selection performance on the PoV change dataset is better than the end-to-end performance, which is expected due to error propagation from previous stages in the pipeline. The proposed LSTM-based models significantly outperform the baselines, and also perform better than the decision tree approaches. The mention selection performance on the PoV dataset is slightly better than on CoNLL. This is because the CoNLL performance is calculated solely based on mention conversions, while PoV performance numbers are based on both mention selection and verb conjugations, which have a relatively higher accuracy. All trained models were tuned on the CoNLL validation data. We also tuned on PoV validation data, however that lead to very similar performance.

Table 4 also shows the precision (P), recall (R), and F1 measure of the End-to-End pipeline for each article. The key factor affecting the performance appears to be the average number of mention candidates per entity MC_e, with values greater than 3.8 leading to a more marked decrease in performance. Figure 2 shows the accuracy of the mention selection model *LSTM: Token + Mention* as a function of the size of the set of candidate entities $S(E)$. Overall, as expected, the mention selection accuracy trends down as the size of $S(E)$ increases.

Table 6 shows the results of 5 ablation tests, demonstrating that all components of the mention selection architecture are beneficial to the final End-to-End performance: the token-level LSTM, the mention-level LSTM, the token-level binary features ϕ_t^b,
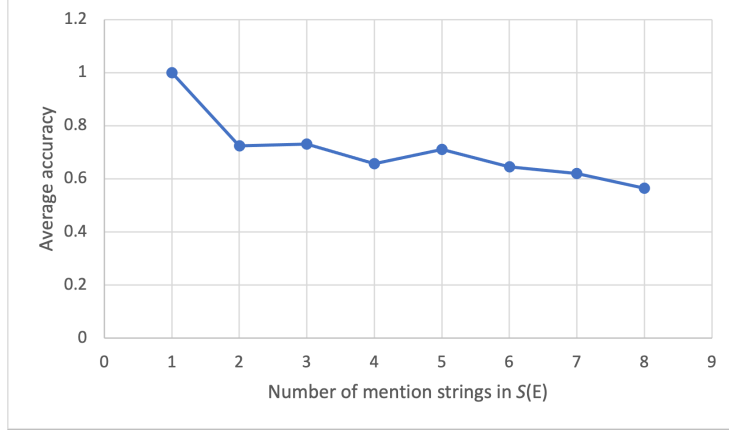
Figure 2: Accuracy of mention selection for different sizes of $S(E)$.

and the overall vector of binary features ϕ^b. We ran statistical test of significance using the *one-tailed paired T-test* over the F_1 results for the 21 documents in the PoV change dataset, i.e. 21 results for each of the systems in the comparison. The following comparisons were highly statistically significant (*p-value* < 0.01):

- *LSTM: Token + Mention* is better than *LSTM: Token*, i.e. the mention-level LSTM is useful.

- *LSTM: Token* + ϕ_t^b is better than *LSTM: Token*, i.e. the token-level binary features are useful.

Furthermore, the following comparison was statistically significant (*p-value* < 0.05):

- *LSTM: Token + Mention* is better than *LSTM: Mention*, i.e. the token-level LSTM is useful too.

Table 7 shows the performance of each component in the system pipeline. For each component, the evaluation is done under the assumption that the preceding components in the pipeline run without errors, e.g. for the verb identification results we assume that the mention detection and coreference resolution, on which verb detection identification depends, are entirely correct.

Sample texts generated by the system are shown in Table 8. Most of these example show that the system is able to make sophisticated mention selection decisions, where

Model	P	R	F$_1$
LSTM: Token + Mention	**71.6**	**73.0**	**72.3**
LSTM: Token + Mention $- \phi^b$	71.3	72.8	72.0
LSTM: Token + Mention $- \phi_t^b$	71.3	72.7	72.0
LSTM: Token	69.7	71.1	70.4
LSTM: Mention	71.0	72.4	71.7
LSTM: Token $+ \phi_t^b$	71.1	72.5	71.8

Table 6: Ablation results (%) for the End-to-End scenario in terms of Precision (P), Recall (R), and F1 score.

Pipeline component	P	R	F1
Coreference resolution	91.1	96.9	93.9
Verb identification	75.2	76.7	75.9
Verb conjugation	100.0	100.0	100.0
Generation of $S(E)$	99.1	99.1	99.1

Table 7: Performance (%) in terms of Precision (P), Recall (R), F1 score for each component in the pipeline.

pronouns alternate with names and noun phrases in order to alleviate potential ambiguity introduced by confounding entities, while also preserving the naturalness of the text. In sentences 1 and 3, for instance, the system learns to not use the same third person pronoun for different entities mentioned in subject position in consecutive clauses.

There are also instances where, even though the system selection is different from the manual annotation, the result still sounds acceptable. In sentence 2, the system preferred "his" over the annotated string "this old man". Upon searching the training data, we found similar possessive constructions, as in "And so this explanation of his is untruthful.", which may explain this mention selection. In sentence 6, even though the system preferred "he" over "his father" for the third mention, it is still an acceptable selection: while it increases the ambiguity, it makes the text sound more natural. For the confounding entity in sentences 7 and 9, the system selected "the first guide" as in the manual annotation, but not "this fellow", preferring to use the name instead. This is probably due to the fact that most of the common nouns used in the CoNLL training data are titles or occupations, such as "the president" or "the correspondent". However, even though the annotated "this fellow" is more coherent with the article's style, the

Focus Entity: **I (Campbell)**, *Confounding Entity*: **George Kraft**

1. When [*he*]$_1$ → [**Kraft**]$_1$ opened up, [*I*]$_2$ → [**Campbell**]$_2$ saw that [*he*]$_1$ was a painter. There was an easel in the middle of [*his*]$_1$ living room with a fresh canvas on it, and there were stunning paintings by [*him*]$_1$ on every wall.

2. George Kraft was only one of [*his*]$_1$ names. The real name of [*this old man*]$_1$ → [**his**]$_1$ was Colonel Iona Potapov.

Focus Entity: **I (Nick)**, *Confounding Entity*: **Nick's father**

3. Even before [*he*]$_2$ became homeless [*I*]$_1$ → [**Nick**]$_1$ had heard whispers, sensed [*he*]$_2$ → [**his father**]$_2$ was circling close, that they were circling each other; like planets unmoored.

4. [*I*]$_1$ → [**Nick**]$_1$ even knew what kind, a Town Taxi, a black and white. In [*my*]$_1$ → [**his**]$_1$ early twenties, after [*I*]$_1$ → [**he**]$_1$ dropped out of college and moved to Boston, [*I*]$_1$ → [**he**]$_1$ would involuntarily check the driver of each that passed, uncertain what it would mean, what [*I*]$_1$ → [**he**]$_1$ would do, if it was [*my*]$_1$ → [**his**]$_1$ father behind the wheel.

5. [*I*]$_1$ → [**Nick**]$_1$ could have given [*him*]$_2$ a key, offered a piece of [*my*]$_1$ → [**his**]$_1$ floor. A futon. A bed. But [*I*]$_1$ → [**he**]$_1$ never did. If [*I*]$_1$ → [**he**]$_1$ let [*him*]$_2$ → [**his father**]$_2$ inside [*I*]$_1$ → [**he**]$_1$ would become [*him*]$_2$.

6. If you asked [*me*]$_1$ → [**him**]$_1$ about [*my*]$_1$ → [**his**]$_1$ father then - the years [*his father*]$_2$ → [**he**]$_1$ lived in a doorway, in a shelter, in an ATM - [*I*]$_1$ → [**Nick**]$_1$ would say, "Dead," [*I*]$_1$ → [**he**]$_1$'d say, "Missing," [*I*]$_1$ → [**he**]$_1$'d say, "I don't know where he is."

Focus Entity: **I (Michael)**, *Confounding Entity*: **Fritz (the guide)**

7. At least on paper, nothing about [*the first guide*]$_1$ [*I*]$_2$ → [**Michael**]$_2$ chose to work with sounds auspicious.

8. [*I*]$_2$ → [**Michael**]$_2$ surprised [*myself*]$_2$ → [**himself**]$_2$ by liking [*Fritz*]$_1$ as much as [*I*]$_2$ → [**he**]$_2$ did.

9. [*I*]$_2$ → [**Michael**]$_2$ could just forget about the whole idea of being anywhere in range of a hospital emergency room. Then there was the fact that while [*I*]$_2$ → [**he**]$_2$ was from a family that had once been reluctant to buy a German car, [*this fellow*]$_1$ → [**Fritz**]$_1$ was a German in [*his*]$_1$ midsixties whose father had served in the army during World War II.

10. [*Fritz*]$_1$ told [*me*]$_2$ → [**Michael**]$_2$ what to expect if [*I*]$_2$ → [**he**]$_2$ *were* → **was** to work with [*him*]$_1$.

Table 8: Sample texts generated by the system. Entity mentions are shown between brackets, indexed by their coreference chain. Highlighted in red are the system outputs that are different from the manual annotations shown in light gray.

system generated mention does decrease the ambiguity.

7.2. Human Evaluation

Evaluating the output of a PoV change system is complicated by the fact that there may be multiple solutions for choosing mention strings that achieve felicitous, non-

ambiguous reference while also maintaining the naturalness of a sentence. Inspired by work in style transfer (Shen et al., 2017; Rao and Tetreault, 2018; Prabhumoye et al., 2018), we conducted human-based evaluations to assess the quality of the *Manual*-ly annotated text vs. the system-generated text in the *Mention Selection* and *End-to-End* scenarios. To obtain high quality annotations from Amazon Mechanical Turk (MTurk) workers, we followed Clark et al. (2018) and only used US-located workers with a HIT approval rate greater than 95% and a number of approved HITs greater than 1,000. Each worker was presented with a paragraph from a story in the PoV dataset, one sentence at a time. For each mention in the sentence, the worker was asked to map it to the corresponding discourse entity, i.e. its referent, by selecting from a predefined list of entities known to appear in the original version of that paragraph. To understand the context of the current sentence, the previous sentences were also shown to the worker. Each worker was asked to annotate three versions of the paragraph: Manual, System: Mention Selection, and System: End-to-End. The three versions were presented in random order.

To estimate referential ambiguity, each worker w was asked to indicate how easy it was to map the mention string m to its referent entity, using a rating $amb(w, m)$ ranging from 0: *Hard* (high ambiguity), to 1: *Moderate* (some ambiguity), to 2: *Easy* (no ambiguity). The workers also had the option of indicating whether the mention could not be mapped to any of the listed entities, using an *Impossible* rating. This can happen when a wrong string is used for that mention, mainly due to mistakes in coreference resolution that propagate in the construction of $S(E)$. Below are examples illustrating ratings of entity disambiguation difficulty, which were provided as part of the instructions to the MTurk workers:

Previous sentences: The door to Reynolds' apartment is also open. Leon walks down the entry-way to the living room, hearing a hyperaccelerated polyphony from a digital synthesizer.

1. **Current sentence**: Evidently, it's *Reynold's* own work; the sounds are modulated in ways undetectable to normal hearing, and even <u>Leon</u> can't discern any pattern to them.

   ◉ Easy ◯Moderate ◯Hard ◯Impossible

   Analysis: The mention string "Leon" clearly refers to Leon. Correspondingly, there is no referential ambiguity.

2. **Current sentence**: Evidently, it's *Reynold's* own work; the sounds are modulated in ways

undetectable to normal hearing, and even <u>he</u> can't discern any pattern to them.

○Easy ◉ Moderate ○Hard ○Impossible

Analysis: The pronoun "he" may refer to either Leon or Reynold. However, according to the context, it is Reynold's own work, therefore it should be Leon who cannot discern the pattern. Correspondingly, there is some referential ambiguity.

3. **Current sentence**: Evidently, it's *his* own work; the sounds are modulated in ways unde-tectable to normal hearing, and even <u>he</u> can't discern any pattern to them.

○Easy ○Moderate ◉ Hard ○Impossible

Analysis: Due to the referential ambiguity of the preceding pronoun "his", we do not know whose work it is, therefore it is hard to determine who "he" refers to. Correspondingly, there is a high level of referential ambiguity for the pronoun "he".

4. **Current sentence**: Evidently, it's *Reynold's* own work; the sounds are modulated in ways undetectable to normal hearing, and even <u>James</u> can't discern any pattern to them.

○Easy ○Moderate ○Hard ◉ Impossible

Analysis: There are only two choices in the entity list: Leon Greco and Reynolds. The name "James" cannot refer to either of them, therefore entity disambiguation is impossible.

Each worker w was also instructed to rate the naturalness $nat(S, w)$ of the current sentence S, where a sentence was considered to sound natural if it was grammatical and not repetitive in terms of the words used for the entity mentions. The naturalness ratings ranged from 0: *Poor*, to 1: *Fair*, to 2: *Good*. Below are examples illustrating ratings of naturalness of sentences which were provided as part of the instructions to the MTurk workers:

1. **Current sentence**: <u>Michael</u> never saw the need for them <u>himself</u>, a little contact or any-thing that even sounded like contact would give <u>him</u> more speed than <u>he</u> could bear.

◉ Good ○Fair ○Poor

Analysis: The sentence is grammatical and appropriately uses the name "Michael" to ground the entity and personal pronouns for subsequent mentions.

2. **Current sentence**: <u>Michael</u> never saw the need for them <u>himself</u>, a little contact or any-thing that even sounded like contact would give <u>Michael</u> more speed than <u>he</u> could bear.

○Good ◉ Fair ○Poor

Analysis: The name "Michael" is unnecessarily used a second time, whereas a pronoun would have been a more natural choice. This makes the sentence sound a bit repetitive.

3. **Current sentence**: <u>Michael</u> never saw the need for them <u>himself</u>, a little contact or anything that even sounded like contact would give <u>Michael</u> more speed than <u>Michael</u> could bear.

   ◯Good ◯Fair ◉ Poor

   Analysis: The name "Michael" is used three time, twice in a row, which makes the sentence sound very repetitive and thus unnatural.

In total, 5 text fragments from the PoV dataset were evaluated by 5 workers each. The number of sentences and mentions in each paragraph are shown in Table 9. In total, a similar number of sentences was used for human evaluations in related tasks, such as (Xu et al., 2012; Clark et al., 2018; Prabhumoye et al., 2018). We also tested if the workers were consistent in their evaluations of the 3 versions, i.e. if previous sentences of any two versions are the same, ideally a worker should give the same ratings whenever the currently evaluated sentence is identical in these two versions. Evaluations that were done at random or that were inconsistent were rejected and the tasks were resubmitted until each text had 5 evaluations.

	T_1	T_2	T_3	T_4	T_5	Total
Sentences	10	8	11	12	10	51
Mentions	21	16	26	30	50	143

Table 9: Number of sentences and mentions in each of the 5 text fragments T_k, and in total.

7.2.1. Human Evaluation Results

For each sentence S evaluated by MTurk workers, we compute a *referential* score $ref(S)$ and a *naturalness* score $nat(S)$. Given the ambiguity $amb(m, w)$ assigned by worker w to mention m, the referential score of sentence S is calculated as follows:

$$ref(S, w) = \frac{1}{|S|} \sum_{m \in S} amb(m, w) \cdot C(m, w) \tag{3}$$

where the correctness $C(m, w) = +1$ if m is mapped by w to the correct entity, and $C(m, w) = -1$ otherwise. If the mapping of m was rated as impossible, the corresponding term in the sum above is set directly to -2. The referential score indicates how successful the mention string was in referring to the correct entity. If the worker
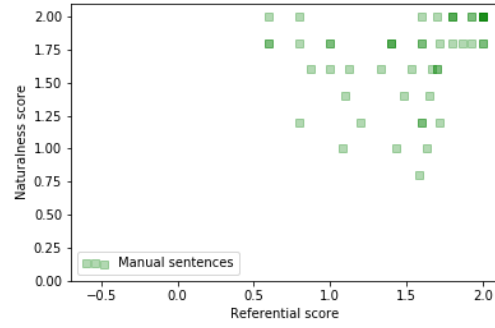
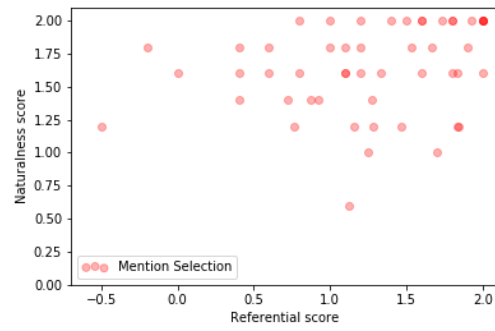Figure 3: Referential and naturalness scores for each evaluated sentence in the Manual version.

Figure 4: Referential and naturalness scores for each evaluated sentence in the Mention Selection version.
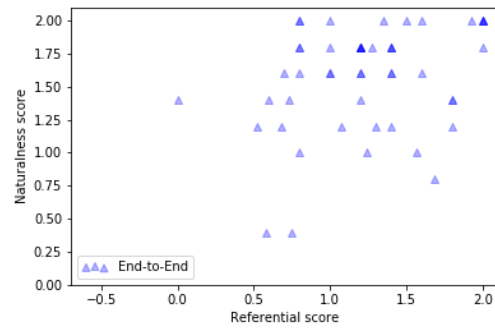

Figure 5: Referential and naturalness scores for each evaluated sentence in the End-to-End version.

mapped the mention to the correct entity, the easier the mapping, the higher the score. Conversely, if the mapping was incorrect, the more confident the worker was, the higher the penalty. Thus, the referential score ranges from -2 for a confident but incorrect entity mapping to $+2$ for a confident and correct mapping. A referential score of -2 is also given to an impossible mention-entity mapping caused by coreference errors. The overall referential and naturalness scores for a sentence are calculated by averaging over the 5 workers:

$$ref(S) \; = \; \frac{1}{5} \sum_{w=1}^{5} ref(S, w) \tag{4}$$

$$nat(S) \; = \; \frac{1}{5} \sum_{w=1}^{5} nat(S, w) \tag{5}$$

Figures 3, 4 and 5 show scatter plots of the naturalness vs. referential scores for all sentences from the three versions. Darker colors indicate multiple sentences with the same scores. Sentences scoring high on both criteria should appear close to the top right corner, corresponding to the maximum possible scores $ref(S) = 2$ and $nat(S) = 2$. In terms of naturalness, the distribution of the system generated sentences (Mention Selection and End-to-End) is comparable to the manually annotated annotations. We attribute this at least in part to the use of contextual BERT embedding, which helps in selecting mention strings that lead to syntactically correct sentences. The system has the tendency to select pronouns over noun phrases, which makes the generated sentences sound natural but may also introduce ambiguity, as evidenced in Figures 4 and 5 where the system referential scores are distributed slighter farther away from the optimal.

The overall referential and naturalness scores of the generated mentions of the 3 versions are shown in Table 10. If we map the $[-2, 2]$ score ranges to the $[0, 100]$ interval so that 0 correspond to the minimum score possible, in terms of referential scores the Mention Selection system obtains 81% of the possible maximum, the End-to-End system obtains 80.75%, compared to 87.25% for our Manual annotations. In terms of naturalness, the Mention Selection system obtains 91% of the possible maximum, the End-to-End system obtains 89.25%, compared to 91.75% for the Manual annotation.

	End-to-End		M. Selection		Manual	
	μ	σ	μ	σ	μ	σ
ref	1.23	0.46	1.24	0.59	1.49	0.42
nat	1.57	0.40	1.64	0.34	1.67	0.32

Table 10: The mean μ and standard deviation σ of the referential scores *ref* and naturalness scores *nat* over all sentences in the human evaluation.

8. Conclusion and Future Work

We introduced the task of changing the narrative perspectives and proposed a pipeline architecture where a mention selection model is trained on data extracted from a corpus annotated with coreference chains. We annotated a benchmark dataset with point of view changes, using text from a wide range of types of narratives. Evaluations on a newly created PoV benchmark dataset showed that the proposed architecture substantially outperforms the baselines and comes close to the manual annotation in terms of referential quality and naturalness. To account for the existence of multiple solutions to the PoV change of a text, we also ran and compared human evaluations of the system output and the manual annotations. To the best of our knowledge, the system presented in this paper is the first to address the problem of changing the narrative perspective of a text. To enable progress in this area, we are making the datasets and the code publicly available. The source code and datasets are publicly available at `https://github.com/chenmike1986/change_pov`.

The approach introduced in this paper has the potential for a wide range of applications, such as the generation of self-diagnosis texts, or text written from a neutral, third-person perspective. The manipulation of narrative perspectives can also be valuable in fiction, where the aim would be to alter the reading experience.

In future work, we plan to extend the approach such that text between mentions is also allowed to change, which is expected to further improve the naturalness of the converted text.

9. Acknowledgements

References

Asimov, I., 1953. Nobody here but –. `https://www.goodreads.com/book/show/24967968-nobody-here-but`. Accessed October, 2020.

Atwood, M., 1998. The Handmaid's Tale. Anchor Books.

Belz, A., Kow, E., Viethen, J., Gatt, A., 2010. Generating Referring Expressions in Context: The GREC Task Evaluation Challenges. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 294–327. URL: `https://doi.org/10.1007/978-3-642-15573-4_15`, doi:10.1007/978-3-642-15573-4_15.

Biss, E., 2009. Notes from No Man's Land. Graywolf Press.

Black, D., 2017. Why can i see my avatar? embodied visual engagement in the third-person video game. Games and Culture 12, 179–199. URL: `https://doi.org/10.1177/1555412015589175`, doi:10.1177/1555412015589175, arXiv:https://doi.org/10.1177/1555412015589175.

Bohman, J., 2000. The importance of the second person: Interpretation, practical knowledge, and normative attitudes, in: Empathy And Agency: The Problem Of Understanding In The Human Sciences. URL: `https://doi.org/10.4324/9780429500831`.

Brunyé, T.T., Ditman, T., Mahoney, C.R., Augustyn, J.S., Taylor, H.A., 2009. When You and I share perspectives: Pronouns modulate perspective taking during narrative comprehension. Psychology Science 20.

Bryson, B., 2010. A Walk in the Woods. Broadway Books.

Campos, R., Jorge, A.M., Jatowt, A., Bhatia, S., Finlayson, M. (Eds.), 2021. Proceedings of the Fourth International Workshop on Narrative Extraction from Texts (Text2Story), Lucca, Italy. URL: `http://text2story21.inesctec.pt/`. To appear.

Center, F.C.M., 2018. Fulton county medical center main page. URL: `http://fcmcpa.org/The-Importance-of-Healthy-Doctor-Patient-Relationships`. accessed October, 2018.

Chiang, T. (Ed.), 1991. Understand. Asimov's Science Fiction.

Choifer, A., 2018. A new understanding of the first-person and third-person perspectives. Philosophical Papers 47, 333–371. URL: `https://doi.org/10.1080/05568641.2018.1450160`, doi:`10.1080/05568641.2018.1450160`, arXiv:`https://doi.org/10.1080/05568641.2018.1450160`.

Chu, J., 2013. The water that falls on you from nowhere. `https://www.tor.com/2013/02/20/the-water-that-falls-on-you-from-nowhere/`. Accessed October, 2020.

Clark, E., Ji, Y., Smith, N.A., 2018. Neural text generation in stories using entity representations as context, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana. pp. 2250–2260. URL: `https://www.aclweb.org/anthology/N18-1204`, doi:`10.18653/v1/N18-1204`.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P., 2011. Natural language processing (almost) from scratch. Journal of Machine Learning Research 12, 2493–2537. URL: `http://jmlr.org/papers/v12/collobert11a.html`.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota. pp. 4171–4186. URL: `https://www.aclweb.org/anthology/N19-1423`, doi:`10.18653/v1/N19-1423`.

Dozat, T., Manning, C.D., 2017. Deep biaffine attention for neural dependency parsing, in: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings, OpenReview.net. URL: `https://openreview.net/forum?id=Hk95PK9le`.

Eisenberg, J.D., Finlayson, M.A., 2016. Automatic identification of narrative diegesis and point of view, in: Proceedings of the 2nd Workshop on Computing News Storylines. URL: `http://www.aclweb.org/anthology/W16-5705`.

Evans, M.T., 2015. Narcissistic abuse - You're damned if you do, damned if you don't. `https://blog.melanietoniaevans.com/narcissistic-abuse-youre-damned-if-you-do-damned-if-you-dont/`. Accessed October, 2020.

Ferreira, T.C., Moussallem, D., Kádár, Á., Wubben, S., Krahmer, E., 2018. Neuralreg: An end-to-end approach to referring expression generation, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1959–1969.

Fludernik, M., 1993. Second person fiction: Narrative "you" as addressee and/or protagonist. AAA: Arbeiten aus Anglistik und Amerikanistik 18, 217–247. URL: `http://www.jstor.org/stable/43023644`.

Flynn, N., 2005. Another Bullshit Night in Suck City. W.W. Norton.

Friedrich, A., Pinkal, M., 2015. Discourse-sensitive automatic identification of generic expressions, in: Proceedings of the 53rd Annual Meeting of the Association for

Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Beijing, China. pp. 1272–1281. URL: `https://www.aclweb.org/anthology/P15-1123`, doi:10.3115/v1/P15-1123.

Fu, Z., Tan, X., Peng, N., Zhao, D., Yan, R., 2018. Style transfer in text: Exploration and evaluation, in: Proceedings of AAAI.

Gardent, C., Shimorina, A., Narayan, S., Perez-Beltrachini, L., 2017. The WebNLG challenge: Generating text from RDF data, in: Proceedings of the 10th International Conference on Natural Language Generation, Association for Computational Linguistics, Santiago de Compostela, Spain. pp. 124–133. URL: `https://www.aclweb.org/anthology/W17-3518`, doi:10.18653/v1/W17-3518.

Gast, N.G., Tomozei, D.C., Le Boudec, J.Y., 2013. Optimal generation and storage scheduling in the presence of renewable forecast uncertainties , 11URL: `http://infoscience.epfl.ch/record/183046`.

Grosz, B.J., Joshi, A.K., Weinstein, S., 1995. Centering: A framework for modeling the local coherence of discourse. Computational Linguistics 21, 203–225. URL: `https://www.aclweb.org/anthology/J95-2003`.

Harrigan, P., Wardrip-Fruin, N., 2007. Second Person: Role-Playing and Story in Games and Playable Media. MIT Press, Cambridge, MA.

Hartung, F., Burke, M., Hagoort, P., Willems, R.M., 2016. Taking perspective: Personal pronouns affect experiential aspects of literary reading. PLOS ONE 11.

Herr, M. (Ed.), 1977. Dispatches. Alfred A. Knopf.

Heylighen, F., Marc Dewaele, J., 1999. Formality of Language: Definition, measurement and behavioral determinants. Technical Report.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation 9, 1735–1780. URL: `http://dx.doi.org/10.1162/neco.1997.9.8.1735`, doi:10.1162/neco.1997.9.8.1735.

Honnibal, M., Montani, I., 2017. Spacy. `https://spacy.io/`.

Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., Weischedel, R., 2006. OntoNotes: The 90% solution, in: Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers, Association for Computational Linguistics, New York City, USA. pp. 57–60. URL: `https://www.aclweb.org/anthology/N06-2015`.

Jaeger, T.F., 2010. Redundancy and reduction: speakers manage syntactic information density. , 23–62URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896231/`.

Jaeger, T.F., Buz, E. (Eds.), 2016. Signal Reduction and Linguistic Encoding. John Wiley & Sons, Inc.

Jhamtani, H., Gangal, V., Hovy, E., Nyberg, E., 2017. Shakespearizing modern language using copy-enriched sequence to sequence models, in: Proceedings of the Workshop on Stylistic Variation, Association for Computational Linguistics, Copenhagen, Denmark. pp. 10–19. URL: `https://www.aclweb.org/anthology/W17-4902`, doi:`10.18653/v1/W17-4902`.

Ji, Y., Tan, C., Martschat, S., Choi, Y., Smith, N.A., 2017. Dynamic entity representations in neural language models, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Copenhagen, Denmark. pp. 1830–1839. URL: `https://www.aclweb.org/anthology/D17-1195`, doi:`10.18653/v1/D17-1195`.

Jorge, A.M., Campos, R., Jatowt, A., Nunes, S., 2019. Information Processing & Management Journal Special Issue on Narrative Extraction from Texts (Text2Story): Preface. Information Processing & Management 56, 1771–1774. URL: `https://www.sciencedirect.com/science/article/pii/S0306457319304455`, doi:`https://doi.org/10.1016/j.ipm.2019.05.004`.

Key, H.S., 2015. My dad tried to kill me with an alligator. `https://www.`

outsideonline.com/1979226/my-dad-tried-kill-me-alligator. Accessed October, 2020.

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization, in: ICLR.

Koul, S., 2016. There's no recipe for growing up. https://www.buzzfeed.com/scaachikoul/looking-for-my-mother-at-the-bottom-of-a-pot. Accessed October, 2020.

Kurosawa, A., 1983. Akira Kurosawa: Something Like an Autobiography. Vintage.

Lee, K., He, L., Lewis, M., Zettlemoyer, L., 2017. End-to-end neural coreference resolution, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Copenhagen, Denmark. pp. 188–197. URL: https://www.aclweb.org/anthology/D17-1018, doi:10.18653/v1/D17-1018.

Lim, S., Reeves, B., 2009. Being in the game: Effects of avatar choice and point of view on psychophysiological responses during play. Media Psychology 12, 348–370. URL: https://doi.org/10.1080/15213260903287242, doi:10.1080/15213260903287242, arXiv:https://doi.org/10.1080/15213260903287242.

van Lissa, C.J., Caracciolo, M., van Duuren, T., van Leuveren, B., 2016. Difficult empathy: The effect of narrative perspective on readers' engagement with a first-person narrator. DIEGESIS 5.

Mani, I., 2013. Computational Modeling of Narrative. Morgan & Claypool Publishers.

Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D., 2014. The Stanford CoreNLP natural language processing toolkit, in: Association for Computational Linguistics (ACL) System Demonstrations, pp. 55–60. URL: http://www.aclweb.org/anthology/P/P14/P14-5010.

Miller, M., 2016. The 37. http://www.joylandmagazine.com/regions/south/37. Accessed October, 2020.

Mitchell, M., Huang, T.H.K., Ferraro, F., Misra, I. (Eds.), 2018. Proceedings of the First Workshop on Storytelling, Association for Computational Linguistics, New Orleans, Louisiana. URL: `https://www.aclweb.org/anthology/W18-1500`, doi:`10.18653/v1/W18-15`.

Modi, A., Titov, I., Demberg, V., Sayeed, A., Pinkal, M., 2017. Modeling semantic expectation: Using script knowledge for referent prediction. Transactions of the Association for Computational Linguistics 5, 31–44. URL: `https://www.aclweb.org/anthology/Q17-1003`, doi:`10.1162/tacl_a_00044`.

Mondimore, F.M., 2006. Bipolar Disorder: A Guide for Patients and Families. John Hopkins University Press.

Morrison, T., 2015. Sweetness. `https://www.newyorker.com/magazine/2015/02/09/sweetness-2`. Accessed October, 2020.

Newell, E., Cheung, J.C., 2018. Constructing a lexicon of relational nouns, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan. URL: `https://www.aclweb.org/anthology/L18-1537`.

Pareti, S., O'Keefe, T., Konstas, I., Curran, J.R., Koprinska, I., 2013. Automatically detecting and attributing indirect quotations, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Seattle, Washington, USA. pp. 989–999. URL: `https://www.aclweb.org/anthology/D13-1101`.

Pauen, M., 2012. The second-person perspective. Inquiry 55, 33–49. URL: `https://doi.org/10.1080/0020174X.2012.643623`, doi:`10.1080/0020174X.2012.643623`, arXiv:`https://doi.org/10.1080/0020174X.2012.643623`.

van Peer, W., Chatman, S. (Eds.), 2001. New Perspectives on Narrative Perspective. The Margins of Literature, SUNY Press.

Platform, B.E., 2018. Parenting advice article. URL: `https://www.betterhelp.com/advice/parenting/`

`what-is-a-toxic-mother-and-how-does-she-affect-relationships/`. accessed October, 2018.

Poesio, M., Stevenson, R., Di Eugenio, B., Hitzeman, J., 2004. Centering: A parametric theory and its instantiations. Computational Linguistics 30, 309–363. URL: `https://www.aclweb.org/anthology/J04-3003`, doi:10.1162/0891201041850911.

Pollan, M., 2018. How to Change Your Mind. Penguin Press.

Possinger, J., 2018. 5 signs of a good lawyer. URL: `https://medium.com/possingerlaw/5-signs-of-a-good-lawyer-b846b6e892e7`. accessed October, 2020.

Prabhumoye, S., Tsvetkov, Y., Salakhutdinov, R., Black, A.W., 2018. Style transfer through back-translation, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia. pp. 866–876. URL: `https://www.aclweb.org/anthology/P18-1080`, doi:10.18653/v1/P18-1080.

Pradhan, S., Moschitti, A., Xue, N., Uryupina, O., Zhang, Y., 2012. CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes, in: Proceedings of the Sixteenth Conference on Computational Natural Language Learning (CoNLL 2012), Jeju, Korea.

Prince, E.F., 1981. Toward a taxonomy of given-new information, in: Cole, P. (Ed.), Syntax and semantics: Vol. 14. Radical Pragmatics. Academic Press, New York, pp. 223–255.

Rao, S., Tetreault, J., 2018. Dear sir or madam, may I introduce the GYAFC dataset: Corpus, benchmarks and metrics for formality style transfer, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana. pp. 129–140. URL: `https://www.aclweb.org/anthology/N18-1012`, doi:10.18653/v1/N18-1012.

Salem, S., Weskott, T., Holler, A., 2017. Does narrative perspective influence readers'
perspective-taking? An empirical study on free indirect discourse, psycho-narration
and first-person narration. Glossa: A Journal of General Linguistics 2.

Samatar, S., 2013. Selkie stories are for losers. `http://strangehorizons.com/`
`fiction/selkie-stories-are-for-losers/`. Accessed October, 2020.

Sartre, J.P., 1969. Nausea. New Directions.

Shen, T., Lei, T., Barzilay, R., Jaakkola, T., 2017. Style transfer from
non-parallel text by cross-alignment, in: Guyon, I., Luxburg, U.V., Ben-
gio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.),
Advances in Neural Information Processing Systems 30. Curran Asso-
ciates, Inc., pp. 6830–6841. URL: `http://papers.nips.cc/paper/`
`7259-style-transfer-from-non-parallel-text-by-cross-alignment.`
`pdf`.

Smith, C.S., 2003. Introduction. Cambridge University Press. Cambridge Studies in
Linguistics, p. 1–4. doi:`10.1017/CBO9780511615108.002`.

Solnit, R., 2013. The Faraway Nearby. Viking.

Strube, M., Hahn, U., 1999. Functional centering: Grounding referential coherence in
information structure. Computational Linguistics 25, 309–344.

Underwood, W., 2009. Examples of Performative Sentences in Presidential Records.
Technical Report.

Ungerleider, S., 2016. I'm a doctor. Preparing you for death is as much a part
of my job as saving lives. `https://www.vox.com/2015/10/19/9554583/`
`doctor-good-death`. Accessed October, 2020.

Vonnegut, K., 1999. Mother Night. Dial Press Trade Paperback.

Walker, M.A., 1993. Initial contexts and shifting centers, in: Proceedings of the Work-
shop on Centering, University of Pennsylvania, Philadelphia.

Weischedel, R., Pradhan, S., Ramshaw, L., et al., 2011. Ontonotes release 4.0, in: Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation. Springer.

Wiebe, J.M., 1994. Tracking point of view in narrative. Computational Linguistics 20, 233–287. URL: `http://www.aclweb.org/anthology/J94-2004`.

Xu, W., Ritter, A., Dolan, B., Grishman, R., Cherry, C., 2012. Paraphrasing for style, in: Proceedings of COLING 2012, The COLING 2012 Organizing Committee, Mumbai, India. pp. 2899–2914. URL: `https://www.aclweb.org/anthology/C12-1177`.